AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1002/2016JD025568

Key Points:
- NPF at five sites across North America exhibits positive 1 day autocorrelation
- NPF frequency, formation and growth rates, survival probabilities, and ultrafine particle concentrations are spatially variable
- Similarities and variability in NPF characteristics can be partly explained by satellite-based measurements of atmospheric composition

Supporting Information:
- Supporting Information S1

Correspondence to:
R. C. Sullivan,
rcs365@cornell.edu

Citation:
Sullivan, R. C., P. Crippa, A. G. Hallar, L. Clarisse, S. Whitburn, M. Van Damme, W. R. Leaitch, J. T. Walker, A. Khlystov, and S. C. Pryor (2016), Using satellite-based measurements to explore spatiotemporal scales and variability of drivers of new particle formation, *J. Geophys. Res. Atmos.*, *121*, doi:10.1002/2016JD025568.

Received 23 JUN 2016
Accepted 22 SEP 2016
Accepted article online 25 SEP 2016

Using satellite-based measurements to explore spatiotemporal scales and variability of drivers of new particle formation

R. C. Sullivan[1], P. Crippa[2], A. G. Hallar[3,4], L. Clarisse[5], S. Whitburn[5], M. Van Damme[5], W. R. Leaitch[6], J. T. Walker[7], A. Khlystov[8], and S. C. Pryor[1,9]

[1]Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, New York, USA, [2]COMET, School of Civil Engineering and Geosciences, Newcastle University, Newcastle upon Tyne, UK, [3]Storm Peak Laboratory, Desert Research Institute, Steamboat Springs, Colorado, USA, [4]Atmospheric Science Department, University of Utah, Salt Lake City, Utah, USA, [5]Spectroscopie de l'Atmosphère, Service de Chimie Quantique et Photophysique, Université Libre de Bruxelles, Bruxelles, Belgium, [6]Environment Canada, Toronto, Ontario, Canada, [7]National Risk Management Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Durham, North Carolina, USA, [8]Desert Research Institute, Reno, Nevada, USA, [9]Pervasive Technology Institute, Indiana University, Bloomington, Indiana, USA

Abstract New particle formation (NPF) can potentially alter regional climate by increasing aerosol particle (hereafter particle) number concentrations and ultimately cloud condensation nuclei. The large scales on which NPF is manifest indicate potential to use satellite-based (inherently spatially averaged) measurements of atmospheric conditions to diagnose the occurrence of NPF and NPF characteristics. We demonstrate the potential for using satellite-based measurements of insolation (UV), trace gas concentrations (sulfur dioxide (SO_2), nitrogen dioxide (NO_2), ammonia (NH_3), formaldehyde (HCHO), and ozone (O_3)), aerosol optical properties (aerosol optical depth (AOD) and Ångström exponent (AE)), and a proxy of biogenic volatile organic compound emissions (leaf area index (LAI) and temperature (*T*)) as predictors for NPF characteristics: formation rates, growth rates, survival probabilities, and ultrafine particle (UFP) concentrations at five locations across North America. NPF at all sites is most frequent in spring, exhibits a one-day autocorrelation, and is associated with low condensational sink (AOD × AE) and HCHO concentrations, and high UV. However, there are important site-to-site variations in NPF frequency and characteristics, and in which of the predictor variables (particularly gas concentrations) significantly contribute to the explanatory power of regression models built to predict those characteristics. This finding may provide a partial explanation for the reported spatial variability in skill of simple generalized nucleation schemes in reproducing observed NPF. In contrast to more simple proxies developed in prior studies (e.g., based on AOD, AE, SO_2, and UV), use of additional predictors (NO_2, NH_3, HCHO, LAI, T, and O_3) increases the explained temporal variance of UFP concentrations at all sites.

1. Introduction and Motivation

New particle formation (NPF) events generate large concentrations of ultrafine particles (UFPs; particle diameter (D_p) < 100 nm), often occur on regional scales, and exhibit high temporal autocorrelation (multiday persistence) [*Hussein et al.*, 2009; *Jeong et al.*, 2010; *Crippa and Pryor*, 2013]. Therefore, NPF may substantially increase the concentration of particles with $D_p \geq 100$ nm and cloud condensation nuclei (CCN) [*Spracklen et al.*, 2008b; *Merikanto et al.*, 2009; *Yu and Luo*, 2009; *Pierce et al.*, 2012, 2014], and thus impact regional climates [*Spracklen et al.*, 2008a; *Paasonen et al.*, 2013]. However, the magnitude of this effect remains uncertain [*Carslaw et al.*, 2013].

Although the precise atmospheric conditions conducive to NPF are not fully understood [*Boy et al.*, 2007] and may vary in space and time [*Kulmala et al.*, 2004; *Yu and Luo*, 2009; *Westervelt et al.*, 2013; *Yu and Hallar*, 2014; *Yu et al.*, 2015], most observational studies are consistent with ternary nucleation involving sulfuric acid (H_2SO_4), water vapor, and some other low-volatility and/or stabilizing condensable species (e.g., oxidation products of biogenic volatile organic compounds (BVOCs) or ammonia (NH_3)) [*Kulmala et al.*, 2000; *Metzger et al.*, 2010; *Sipilä et al.*, 2010; *Zhang et al.*, 2010; *Kirkby et al.*, 2011; *Pryor et al.*, 2011; *Riccobono et al.*, 2014]. Further, the intensity and probability of NPF appear to be positively associated with high insolation and a reduction of the condensational sink (CS) (and thus competition for semivolatile species) [*O'Dowd et al.*, 2002; *Zhang et al.*, 2004a; *Sipilä et al.*, 2010; *Pryor et al.*, 2011; *Almeida et al.*, 2013; *Pierce et al.*, 2014].

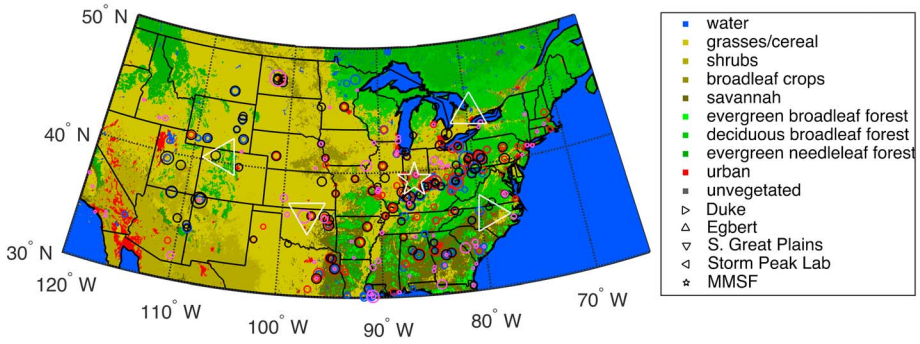

Figure 1. Locations at which the particle size distribution measurements analyzed herein were taken (see Table 1 for details). Background denotes the land cover classification from the MODIS combined Terra and Aqua data set using the LAI/FPAR scheme (type 3) for 2012 [*Land Processes Distributed Active Archive Center [LP DAAC]*, 2014]. The overlaid circles are the 100 largest point source emissions for $PM_{2.5}$ (blue), SO_2 (red), NO_x (black), and NH_3 (magenta) from the EPA 2011 National Emissions Inventory [*U.S. Environmental Protection Agency*, 2011], arbitrarily scaled as a fraction of the largest single emission source for each pollutant. Note, point sources do not clearly reflect the spatial patterns of NH_3 emissions from animal and fertilizer sources. Annual mean particle and trace gas concentrations from the satellite measurements used here are shown in Figure S1.

In situ studies indicate that although initial growth of the recently nucleated particles is largely due to coagulation and condensation of the nucleating gases, growth beyond diameters of tens of nanometers may exhibit an increased contribution from condensation of semivolatile, low-volatility, and extremely low volatility organic gases [*Zhang et al*., 2004a; *Smith et al*., 2005, 2008; *Knol et al*., 2009; *Pryor et al*., 2011; *Pierce et al*., 2012; *Yu and Hallar*, 2014; *Jokinen et al*., 2015]. These commonalities coupled with the occurrence of NPF on regional scales has led to suggestions that nucleation mode and UFP concentrations can be predicted by using satellite-based, and thus spatially averaged, atmospheric properties such as accumulation mode particle properties (aerosol optical depth (AOD) and Ångström exponent (AE)), trace gas concentrations (sulfur dioxide (SO_2) and nitrogen dioxide (NO_2)), and ultraviolet irradiance (UV) [*Kulmala et al*., 2011; *Crippa et al*., 2013; *Sundström et al*., 2015]. More recent work at a single site in the Midwestern USA indicated potential to diagnose not only the total UFP concentrations from remote sensing observations, but also the probability of NPF and descriptors of NPF events: particle formation rates (J_n, where "n" is determined by the minimum detectable particle diameter for each instrument), growth rates (GR), and survival probabilities (SP) [*Sullivan and Pryor*, 2016].

Based on this prior research, we postulate that satellite-derived observations of key properties known, or theorized, to determine the frequency and temporal persistence of NPF, J_n, GR, SP, and UFP concentrations ($N_{n-100\,nm}$) may also be used diagnostically to explain some of the observed spatial variability in these characteristics of NPF [*Pierce et al*., 2014; *Qi et al*., 2015; *Rose et al*., 2015] and reported variations in the closure between models based on simplified nucleation schemes (with fixed nucleation rate coefficients) and observations [*Spracklen et al*., 2008b; *Zhang et al*., 2010; *Lee et al*., 2013]. Given the likely impact of NPF on climate, and substantial uncertainty and model-to-model variability in simulating NPF, improved treatment of NPF in global models is critical for improved understanding of aerosol-climate impacts. Herein we use in situ particle size distribution (PSD) measurements from five sites distributed across North America (Figure 1 and Table 1) to address the following research questions:

1. Can satellite-based measurements of parameters known to be important to NPF and subsequent growth (or proxies for those variables) be used to explain site-to-site variations in NPF frequency and 1 day autocorrelation? For example, is there a relationship between the inherent spatial scales of coherence of the satellite-based measurements of the drivers of NPF and the 1 day autocorrelation in NPF occurrence, and if so which one of the drivers appears to limit the persistence of NPF?
2. Do satellite-based observations offer insights into the causes of variability in NPF characteristics (probability of NPF, J_n, GR, SP, and $N_{n-100\,nm}$) at the five sites?
3. Do proxy algorithms wherein the predictands are the NPF characteristics at the five sites and the predictors are drawn from the suite of remote sensing parameters exhibit commonalities in terms of the most important predictors and the variance explained?

Table 1. Description of Particle Size Distribution Measurement Sites and Instrumentation[a]

Site	Location	Elevation (m)	Dates	Instrument[b]	D_p Range	Bins	Reference
Duke Forest, NC	35.98 N, 79.09 W	179	11/2005 to 09/2007	SMPS	7–289 nm	103	*Pillai et al.* [2013]
Egbert, ON, CAN	44.23 N, 79.78 W	251	05/2007 to 05/2008	SMPS	11–398 nm	26	*Pierce et al.* [2014]
Southern Great Plains, OK (SGP)	36.61 N, 97.49 W	312	01/2010 to 11/2014	TDMA	12 nm–15 μm	207	*Ackerman and Stokes* [2003][c]
Storm Peak Laboratory, CO (SPL)	40.46 N, 106.74 W	3210	03/2012 to 07/2014	SMPS	9–346 nm	104	*Hallar et al.* [2011, 2016]
Morgan Monroe State Forest, IN (MMSFa)	39.32 N, 86.42 W	275	01/2007 to 03/2009	SMPS	6–100 nm	81	*Pryor et al.* [2010]
Morgan Monroe State Forest, IN (MMSFb)	39.32 N, 86.42 W	275	03/2012 to 12/2013	FMPS	6–523 nm	32	*Pryor et al.* [2014]

[a]Locations of the sites relative to land use and point source emissions are given in Figure 1.
[b]SMPS: Scanning Mobility Particle Sizer, FMPS: Fast Mobility Particle Sizer, TDMA: Tandem Differential Mobility Analyzer.
[c]These data are acquired as part of the U.S. Department of Energy's Atmospheric Radiation Measurement program, and to our knowledge have not been published elsewhere; the program and the site are described in this reference.

4. Does a proxy model of UFP concentrations using a larger suite of predictors exhibit more explanatory power than the prior satellite-based proxies, which have employed AOD and AE, SO_2 (or NO_2), and UV?

Additionally, given in situ PSD measurements are time-consuming and expensive, and thus are typically made for limited time periods, we perform a statistical analysis to quantify how using a limited sample of environmental conditions impacts the generalizability of inferences drawn from data collected during time-limited campaigns.

2. Methods

2.1. Particle Size Distribution Data

The PSD measurements used herein all have durations of a year or more and derive from locations distributed across North America with different land use and proximity to major point source emissions (Table 1 and Figure 1). While all PSD measurements are taken at/near the surface, the high elevation of Storm Peak Laboratory (SPL) renders it representative of free tropospheric air on a near daily basis [*Yu and Hallar*, 2014]. Prior to presenting the derived NPF descriptors it is important to note that the individual sites used different instrumentation (and thus have different minimum D_p detection limits; Table 1) and sampling protocols that may confound intercomparison across the five sites. For example, archiving of data from Southern Great Plains (SGP) at 30 min resolution reduces the confidence in the calculated NPF metrics for that site. However, these are the longest records of UFP PSD measurements currently available for North America. Further, analysis of data collected by using two different instruments (a Fast Mobility Particle Sizer (FMPS) and a Scanning Mobility Particle Sizer (SMPS)) at Morgan Monroe State Forest (MMSF) can be used to partly evaluate the impact of instrumentation versus spatial variability in determining NPF characteristics and drivers.

2.2. Remote Sensing Measurements

Once daily satellite-based observations used herein as predictors of NPF and their associated uncertainties are summarized in Table 2. The justification for the selection of these variables is as follows:

1. The cross product of aerosol optical depth (AOD) and Ångström exponent (AE) is used as a proxy for CS following *Crippa et al.* [2013] and is anticipated to be negatively associated with NPF occurrence, J_n, GR, and SP.
2. SO_2 is used as a proxy for H_2SO_4 following *Crippa et al.* [2013], *Kulmala et al.* [2011], and *Sundström et al.* [2015], although it must be noted that the retrievals exhibit a low signal-to-noise ratio, except near large emissions [*Krotkov et al.*, 2008; *Fioletov et al.*, 2011], and many negative concentrations are reported in the Ozone Monitoring Instrument (OMI) SO_2 product. Although nucleation rates may be $\propto [SO_2]^n$ [*Kuang et al.*, 2008] or dependent on H_2SO_4 production ($\propto SO_2 \times UV$), we do not include exponential or compound predictor variables due to the low sensitivity of satellite-based measurements of SO_2 and to avoid overfitting the regression models. Sensitivity analyses indicate that inclusion of previously used compound variables (e.g., $SO_2 \times UV/CS$ [*Kulmala et al.*, 2011; *Sundström et al.*, 2015]) does not increase the variance explanation of nearly all of the regression models (mean improvement across sites and metrics $< 1\%$). SO_2 is anticipated to be positively associated with NPF occurrence, J_n, GR, and SP.

Journal of Geophysical Research: Atmospheres

10.1002/2016JD025568

Table 2. Description of Daily (1 in 8 Days for LAI) Satellite-Based Measurements Used Herein

Predictor	Satellite/Instrument	Overpass (LST)	Available Beginning (% Missing[b]: total, ≤ 2 days)	Version	Resolution (at Nadir)	Uncertainty/Accuracy[c]	Proxy for	Citation
"Dark target" Aerosol optical depth at 550 nm (AOD)	Terra/MODIS	1030	2000(49 %, 71%)	Collection 6	10 km	$\pm 0.05 \pm 0.15 \times$ AOD	AOD × AE ∝ Condensational sink	*Levy et al*. [2013]
Ångström exponent 470–660 nm (AE)	Terra/MODIS	1030	2000(49%, 71%)	Collection 6	10 km	±0.4[d]		*Levy et al*. [2013]
Sulfur dioxide (SO_2; DU)	Aura/OMI	1345	2004(49%, 75%)	Version 3	13 km × 24 km	Greatest of 1.1 DU, 50%	H_2SO_4 production	*Chance* [2002] and *Brinksma et al*. [2003]
Local solar noon spectral irradiance at 310 nm (UV; mW m^{-2} nm^{-1})	Aura/OMI	1345	2004(20%, 99%)	Version 3	13 km × 24 km	10%	Production of oxidants	*Chance* [2002] and *Brinksma et al*. [2003]
Nitrogen Dioxide (NO_2; molec. cm^{-2})	Aura/OMI	1345	2004(43%, 79%)	Version 3	13 km × 24 km	2×10^{14} molec. cm^{-2} (30%) background (polluted)[e]	Anthropogenic emissions	[*Chance* [2002] and *Brinksma et al*. [2003]
Ammonia (NH_3; molec. cm^{-2})	MetOp/IASI	930[a]	2008(58%, 63%)	NN Version 1	12 km	<100% or $< 5 \times 10^{15}$ molec. cm^{-2}	Ternary nucleation	*Whitburn et al*. [2016]
Formaldehyde (HCHO; molec. cm^{-2})	Aura/OMI	1345	2004(44%, 78%)	Version 3	13 km × 24 km	35%	Production of low-volatility vapors from BVOC	*Chance* [2002] and *Brinksma et al*. [2003]
Leaf area index (LAI; m^2 m^{-2})	Terra and Aqua/MODIS	1030 and 1330	2002	Version 5	1 km	1 m^2 m^{-2}	LAI × T ∝ BVOC emissions	*Fang et al*. [2012]
Daytime land surface temperature (T; K)	Terra and Aqua/MODIS	1030 and 1330	2002	Version 5	1 °	1 K		*Wan* [2008]
Ozone (O_3; DU)	Aura/OMI	1345	2004(43%, 79%)	Version 3	13 km × 24 km	Greatest of 10 DU, 5%	Oxidant and stagnation	*Chance* [2002]; *Brinksma et al*. [2003]

[a]09:30 is the overpass time at the equator. Only AM retrievals are used.
[b]Percentage of missing days (averaged across the five sites and entire satellite observation period), and percentage of missing days with a duration of less than or equal to two consecutive days. MODIS LAI is an 8 day composite product, and thus, missing days are not shown for LAI or *T*.
[c]OMI accuracy: "root sum of the square of all errors, including forward model, inverse model, and instrument errors" [*Brinksma et al*., 2003].
[d]MODIS AE is typically bimodal in nature and thus uncertainty is ambiguous [*Levy et al*., 2010].
[e]When averaged to 26 km × 48 km [*Brinksma et al*., 2003].

3. UV is used as a proxy for photochemical production of oxidants (e.g., the hydroxyl radical (OH)) and thus oxidation of SO_2 to H_2SO_4 and BVOCs to low-volatility products following *Crippa et al*. [2013], *Kulmala et al*. [2011], and *Sundström et al*. [2015], and is anticipated to be positively associated with NPF occurrence, J_n, GR, and SP.
4. NO_2 is used as a proxy for air masses influenced by anthropogenic emissions (including primary emitted particles and condensable vapors) [*Russell et al*., 2012], following *Kulmala et al*. [2011] and *Sundström et al*. [2015]. Additionally, satellite-based measurements of NO_2 have been observed to correlate better with in situ SO_2 measurements than satellite-based retrievals of SO_2 [*Sundström et al*., 2015]. NO_2 is anticipated to be negatively associated with NPF occurrence and SP due to its association with primary particle emissions and thus increased CS, but positively associated with J_n and GR due to increased precursor concentrations.
5. NH_3 may play a role in enhancing NPF by acting as a stabilizing base for nucleating clusters, and is used here following *Crippa et al*. [2013] (daily NH_3 estimates are used herein, versus seasonal averages in *Crippa et al*. [2013]), and is anticipated to be positively associated with NPF occurrence, J_n, GR, and SP.
6. Formaldehyde (HCHO) is a product of oxidation of VOCs and one of the few organic species retrievable from satellite-based measurements. It is used as a proxy for the abundance of low-volatility VOCs [*Chance et al*., 2000; *Henze and Seinfeld*, 2006]. The dominant source of HCHO estimated from satellite-based measurements appears to be the oxidation of isoprene because of the short lifetime of isoprene and production of HCHO in the initial oxidation steps, while most anthropogenic VOC emissions require more oxidation steps prior to HCHO formation and thus are diluted prior to HCHO production [*Millet et al*., 2008]. Organics play a role in NPF and/or growth of newly formed particles at least in some environments [*O'Dowd et al*., 2002; *Zhang et al*., 2004a, 2004b; *Henze and Seinfeld*, 2006; *Metzger et al*., 2010; *Paasonen et al*., 2010; *Pryor et al*., 2011; *Pierce et al*., 2012; *Riipinen et al*., 2012; *Kulmala et al*., 2013], but uncertainty remains regarding whether isoprene products contribute to or suppress NPF [*Surratt et al*., 2006; *Kiendler-Scharr et al*., 2009]. Thus, the expected association between HCHO and NPF occurrence is uncertain, but it is anticipated to be positively associated with J_n, GR, and SP.
7. The cross product of leaf area index (LAI) and skin temperature (T) is used as an additional proxy of BVOC emissions [*Guenther et al*., 1993] and is anticipated to be positively associated with NPF occurrence, J_n, GR, and SP.
8. Ozone (O_3) is both a key atmospheric oxidant [*Helmig*, 1997; *Seinfeld and Pandis*, 2006] and a proxy for atmospheric stagnation [*Valente et al*., 1998]. Total column O_3 as retrieved from the Ozone Monitoring Instrument (OMI) is naturally dominated by stratospheric concentrations, but the presence of a temporal mode of variance at synoptic time scales (see below) indicates that these measurements are also responsive to tropospheric variability. O_3 is anticipated to be positively associated with NPF occurrence, J_n, GR, and SP.

Prior research has demonstrated the potential for using satellite-based measurements as proxies for daily ultrafine or nucleation mode particle concentrations [*Kulmala et al*., 2011; *Crippa et al*., 2013; *Sundström et al*., 2015] using AOD and AE, SO_2, and UV as predictors. We use a proxy based solely on variables 1–3 ("simple model") as a benchmark against which to evaluate whether a model including additional predictor variables: NO_2, NH_3, HCHO, LAI × T, and O_3 ("full model") exhibits improved performance.

All observations as obtained from the respective retrieval teams are subject to the following postprocessing:

1. For the spectral and spatial correlation analyses, spatially consistent time series are required. Thus, the remotely sensed measurements are spatially averaged to a 0.5° × 0.5° grid. This resolution was selected to remove some noise through spatial averaging, without removing important mesoscale variability [*Anderson et al*., 2003]. Due to the lower temporal resolution of the Moderate Resolution Imaging Spectroradiometer (MODIS) LAI measurements (1 in 8 days), LAI × T is excluded from these analyses. For all remaining analyses (Wilcoxon rank sum test, regression trees, and multiple linear regression), all valid retrievals within 100 km of each PSD measurement site are averaged for each observation day to reduce noise, particularly in the trace gas measurements [*Krotkov et al*., 2008; *Fioletov et al*., 2011], and to enhance data availability for these predictands.
2. For all variables, days without valid measurements are filled using a weighted mean of the nearest preceding and succeeding measurement days, which will likely reduce the explanatory power of the regression models built herein. The percentages of missing data are given in Table 2.

3. AE is calculated from AOD at 470 and 660 nm after spatial averaging using the Ångström power law [*Ångström*, 1964].
4. NH_3 measurement availability begins in 2008 (Table 2), and thus, they are only available for the full duration of PSD measurements at SGP, SPL, and MMSFb and for a portion of the PSD measurements at Egbert and MMSFa (Table 1), but many days do not have coincident measurements (e.g., all of 2007) and are filled with a mean NH_3 value. There are no coincident NH_3 and PSD measurements at Duke, and thus, NH_3 is excluded from the regression analysis at this site. Satellite-based measurements of NH_3 can have a high associated uncertainty related to unfavorable atmospheric conditions and/or low NH_3 abundances [*Whitburn et al*., 2016]. Thus, only NH_3 pixels with relative uncertainty of $<$100% of the retrieved concentration or an absolute error of $<5\times10^{15}$ molec. cm^{-2} are used.
5. All OMI pixels impacted by the row anomalies [*Ozone Monitoring Instrument Team*, 2012] are treated as missing data.
6. All OMI and Infrared Atmospheric Sounding Interferometer (IASI) retrievals are filtered using a cloud screen to remove retrieval with cloud fractions of $>$0.3 [*Fioletov et al*., 2011; *McLinden et al*., 2014; *Vinken et al*., 2014]. The MODIS aerosol retrieval algorithm filters out cloud contaminated pixels prior to averaging spectral reflectances and deriving spectral AOD [*Levy et al*., 2013], and therefore, no additional cloud screening is applied here.
7. Prior to regression analyses, NO_2 and HCHO are log transformed to more closely approximate Gaussian distributions and all predictors are converted to standard normal scores. Therefore, any systematic bias in satellite retrievals should not impact the analyses as only relative concentrations are considered. Random errors in the retrievals will propagate through the analyses and are expected to reduce the association between the regression predictors and predictands, but not sign or slope of the relationship.

NPF at the measurement sites typically begin in the morning hours (~9:00–11:00 local standard time (LST); Figure 2c), and event metrics are calculated for the subsequent 3 h, thus are typically centered on the satellite overpass times (9:30–13:45 LST; Table 2). It is noted that the once daily measurements cannot characterize diurnal variability of the predictor variables and may thus reduce the predictive skill of the models built from them, particularly when events do not occur near the satellite overpass (e.g., at SGP). Further, satellite-based measurements are columnar measurements and may not fully characterize near-surface conditions [e.g., *van Donkelaar et al*., 2006, 2013; *Lamsal et al*., 2008; *Sundström et al*., 2015] and may reduce the explanatory power of the regression models built herein.

2.3. Event Classification and Characteristics, and Statistical Methods

To quantify similarities and differences in NPF frequency, persistence, and seasonality across North America, an automated methodology is applied to each of the PSD data sets to identify event occurrence and estimate J_n, GR, and SP (an earlier version of the approach was described in *Sullivan and Pryor* [2016]). In brief, a NPF event is reported, and included in the analysis, if:

1. The minimum nucleation mode geometric mean diameter ($<$30 nm; $D_{g\mathrm{Nuc}}$) occurs within 10 h of the peak nucleation mode number concentration and while the difference in the geometric mean diameter for $D_p < 100$ nm (D_g) and $D_{g\mathrm{Nuc}}$ is less than or equal to 10 nm.
2. The r^2 of the regression fit for the GR calculation (tracking $D_{g\mathrm{Nuc}}$ from event start to +3 h) is $\geq$0.5.
3. And the event metrics can be reasonably calculated (e.g., GR and $J_n > 0$; equations (1)–(6)):

$$J_n = \frac{dN_{\mathrm{nuc}}}{dt} + F_{\mathrm{coag}} + F_{\mathrm{growth}} \ldots \tag{1}$$

$$\text{where } F_{\mathrm{coag}} = \sum_{i\ =\ \min(D_p)}^{i\ =\ 30\ \mathrm{nm}} \left[\frac{1}{2}K_{ii}N_i + \sum_{j\ =\ i\ +\ \Delta D_p}^{j\ =\ \max(D_p)} \left(K_{ij}N_j\right)\right] \ldots \tag{2}$$

$$\text{and } F_{\mathrm{growth}} = \sum_{i\ =\ \min(D_p)}^{i\ =\ 30\ \mathrm{nm}} \left[\frac{N_i}{30\ \mathrm{nm} - D_{pi}} * \mathrm{GR}\right] \ldots \tag{3}$$

$$\mathrm{SP} = \prod_{i\ =\ \min(D_p)}^{i\ =\ 100\ \mathrm{nm} - \Delta D_p} \exp\left(-\frac{\tau^{\mathrm{growth}}_{D_{pi}, D_{pi}\ +\ \Delta D_p}}{\tau^{\mathrm{coag}}_{Dp_i}}\right) \ldots \tag{4}$$

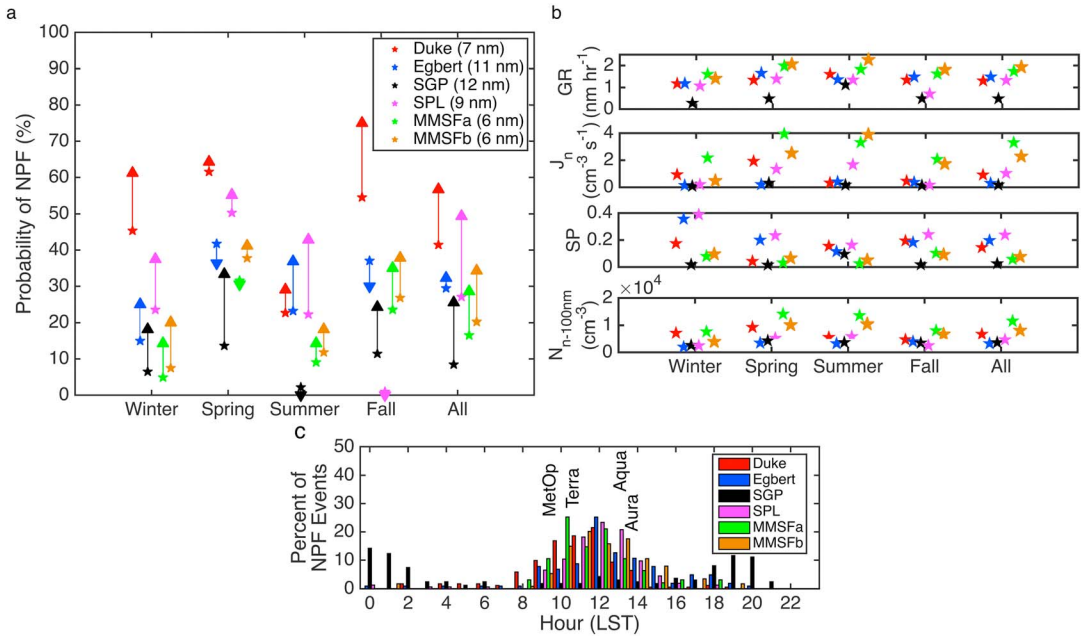

Figure 2. (a) Probability of a NPF event occurring (p(1); stars) and the probability of an event occurring given that an event occurred on the prior day (p(1|1); terminal point of arrows) by season. The instrument's minimum particle D_p measured at each site is given in parenthesis in the legend. (b) Event metrics (growth rates (GR), formation rate (J_n, where n is determined by the minimum detectable particle diameter for each instrument), and survival probabilities (SP)) and daily mean ultrafine particle concentration ($D_p < 100$ nm; $N_{n-100nm}$,) by season and location. (c) Percentage of NPF event days that begin in each hour of the day by measurement site. Also shown are the approximate overpass times for the MetOp (930; NH_3), Terra (1030; AOD × AE, LAI × T), Aqua (1330; LAI × T), and Aura (1345; SO_2, UV, NO_2, HCHO) satellites (ordinate position selected for visibility). Calculations of the event metrics values are described in section 2.3.

$$\text{where } \tau^{\text{growth}}_{D_{pi},\,D_{pi}+\Delta D_p} = \frac{\left(D_{pi}+\Delta D_p - D_{pi}\right)}{\text{GR}} \ldots \tag{5}$$

$$\text{and } \tau^{\text{coag}}_{D_{pi}} = \frac{1}{\frac{1}{2}K_{ii}N_i + \sum_{j=i+\Delta D_p}^{\max(D_p)} K_{ij}N_j} \ldots \tag{6}$$

where K is the Fuch's form Brownian coagulation coefficient, N_{nuc} is the number concentration of particles ≤30 nm, N_i and D_{pi} are the number concentration and median diameter of bin i, ΔD_p is the bin width, and J_n is averaged from event start to +3 h.

The classification algorithm is thus designed to capture unambiguous "A"-type (appearance of nucleation mode particles, followed by clear, sustained growth) NPF events and will classify all days not meeting all of the above criteria as a "nonevent" day even though new particles may be forming (e.g., "B"- or "C"-type events (no appearance of particles in the smallest diameters measured or lack of clear, sustained growth) [*Pryor et al.*, 2010]), thereby reducing NPF frequencies relative to a subjective approach.

Prior research indicates that the dominant NPF mechanism may vary seasonally [*Yu and Hallar*, 2014; *Yu et al.*, 2015]. Thus, the analyses described below are conducted by climatological season, or for leaf-active (defined here as LAI × $T \geq$ 50th percentile) and leaf-dormant (LAI × $T <$ 50th percentile) periods.

The persistence of NPF events is characterized by using the conditional probability of events with a lag of 1 day (i.e., p(1|1)) relative to the probability of an event on any day (p(1)), thus:

1. p(1) ≈ p(1|1): No autocorrelation
2. p(1) < p(1|1): Positive autocorrelation
3. p(1) > p(1|1): Negative autocorrelation

Once the sites have been characterized in terms of the NPF occurrence, intensity, and persistence, we then seek to determine if inherent scales in the NPF predictors (satellite observations) around each site can be used to diagnose and explain the observed consistencies and site-to-site differences in NPF frequency and

event characteristics. To identify the dominant temporal scales of variability in the satellite-based predictors of NPF and the spatial scale on which they exhibit coherence, the time series of remotely sensed parameters at each measurement site were subject to a fast Fourier transformation and used to compute power spectra. The spatial coherence is defined as the distance from the PSD measurement site at which the mean correlation coefficient (Pearson's *r*) between the time series of the predictors at that site and each surrounding grid cell drops below an arbitrary threshold of 0.3.

Finally, we focus on assessing the potential to extend satellite-based proxies of UFP concentrations by expanding both the number of predictor variables used (i.e., the suite of satellite observations included) and the range of characteristics of NPF events considered. We begin by applying the nonparametric Wilcoxon rank sum test to quantify whether the remote sensing predictors exhibit different values on NPF event versus nonevent days at each site. This tests the null hypothesis that it is equally probable that a given observation from one sample is either greater than or less than a given observation from a second sample (different populations across the range of observations, not solely mean or median). We then build regression trees [*Hyvönen et al*., 2005] to recursively partition the predictors based on the occurrence (or not) of NPF. In this way we can determine which predictors (and predictor interactions) are most important in terms of predicting whether an individual day as described using the remote sensing variables will be characterized by NPF or not. The predictors that are most important lie closer to the root node and can be used to interpret how the dependence of NPF on a given predictor variable is conditional on other predictor variables. Finally, multiple linear regression models (equation (7)) are fit in which the predictands are J_n, GR, SP, and $N_{n-100\,\mathrm{nm}}$ and the predictors are all of the remote sensing variables, and the variance explanation are compared with those from a smaller suite of previously used predictors (i.e., AOD × AE, SO_2, and UV [*Kulmala et al*., 2011; *Crippa et al*., 2013; *Sundström et al*., 2015]):

$$y_j = \beta_1 x_{1,j} + \cdots + \beta_i x_{i,j} + \text{constant}\ldots \tag{7}$$

where y_j is the predictand, β_i is the coefficient weighting, and $x_{i,j}$ is the standard normal score of the predictor variable, "*i*", on day, "*j*".

The predictor variable coefficient weights (β_i) are used to diagnose which predictors control each event characteristic and the degree to which they differ among the measurement sites. The results of this analysis are interpreted cautiously because multiple linear regression assumes a linear relationship between the predictors and predictands, multivariate normality, and no multicollinearity among the predictors. To investigate the impact of using finite temporal sample, the multiple linear regression is conducted as a Monte Carlo experiment (1000 iterations), in which we subsample the PSD data sets to train the regression models using a *k*-fold (*k* = five folds) cross validation with 20% of the data withheld from the training model. This is designed to quantify how model skill and coefficient weightings depend on precise time period of field measurements.

3. Results

3.1. NPF Characteristics at the Five Sites

NPF is frequently observed at all five sites with highest NPF frequency, total sub-100 nm particle concentrations ($N_{n-100\,\mathrm{nm}}$), and highest GR in spring, with a secondary peak in NPF frequency in fall (except at SPL, where prior analyses have indicated a secondary fall peak [*Hallar et al*., 2016]) (Figure 2). The discrepancy with prior research at SPL may be due to the lower data availability in fall (due to limited site access), and that while nucleation mode particle formation is observed it is not frequently followed by clear, sustained growth to larger particle sizes (requisite for classification of an event day). Similarly, the low NPF frequency in summer at SGP may be due to missing data in the summer of 2011 and/or the lower temporal frequency (30 min^{-1}) of the PSD measurements at the site resulting in fewer days meeting the strict criteria for an event day.

All sites exhibit an overall positive 1 day autocorrelation of NPF ($p(1) < p(1|1)$) indicating a higher probability of an event if one occurred on the prior day, although this is not observed for all seasons (Figure 2a). Despite these commonalities, there are also differences in NPF characteristics among the five sites. For example, NPF event frequency is substantially higher at Duke than the other sites (Figure 2a). Also, seasonally averaged GR,

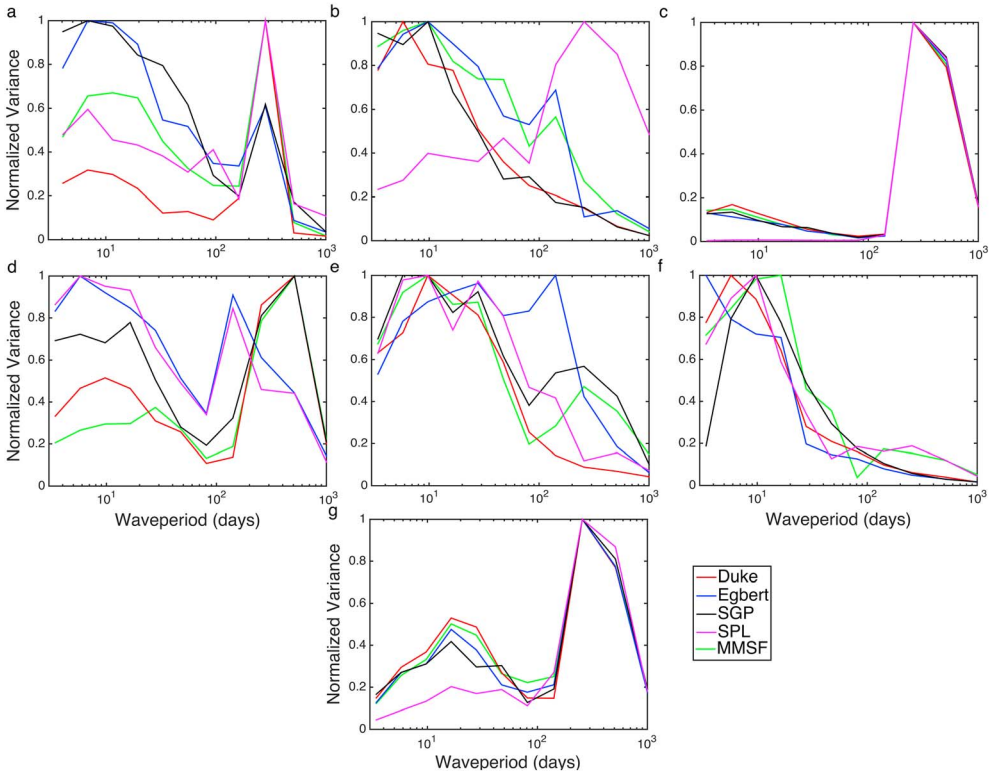

Figure 3. Normalized variance of once daily (a) AOD × AE, (b) SO_2, (c) UV, (d) NO_2, (e) NH_3, (f) HCHO, and (g) O_3 for the grid cells containing each PSD measurement site (indicated by different colors). The variance at each waveperiod is normalized by the maximum variance at any frequency (waveperiod), and the power spectra have been smoothed to emphasize the primary modes of variability. For this analysis the entire satellite measurement period is used. The OMI measurements (SO_2, UV, NO_2, HCHO, and O_3) are from 2004 to 2014, MODIS measurements (AOD × AE) are from 2000 to 2014, and IASI measurements (NH_3) are from 2008 to 2014. Due to low temporal resolution of MODIS LAI (1 in 8 days), LAI × T is omitted.

particle formation rates (J_n), and $N_{n-100\,nm}$ are highest at MMSF in both periods (MMSFa and MMSFb), suggesting that the impact of differences in instrumentation during the two sampling periods on NPF metrics is modest compared to the spatial variability. The observation that GR, J_n, and $N_{n-100\,nm}$ concentrations are highest at MMSF may be due, at least in part, to the lowest min(D_p) at this site, but given the min(D_p) at Duke differed from that at MMSF by only 1 nm, some of the site-to-site variability in these metrics may also reflect spatial variability in NPF events.

3.2. Spatiotemporal Scales of the Predictors

The proposed satellite predictors of NPF exhibit similar dominant scales of temporal variability at all sites (Figure 3). SO_2 exhibits highest variance at synoptic time scales (3–10 days) at all sites except SPL, where the variance is focused on the annual mode. The dominance of the annual mode at SPL may reflect reduced upwind power plant emissions [*Mast et al*., 2005] or the fact that this site is frequently in the free troposphere, while the SO_2 product is designed to represent SO_2 in the planetary boundary layer (Figure 3b) [*Krotkov et al*., 2006; *Fioletov et al*., 2011]. Despite large spatial gradients in isoprene emissions and HCHO concentrations across the study area [*Millet et al*., 2006, 2008], HCHO variability is dominated by the synoptic scale at all sites (Figure 3f), while NO_2 concentrations exhibit high variance on both synoptic and seasonal time scales (Figure 3d). O_3 concentrations and UV are naturally dominated by the annual cycle (Figures 3c and 3g) and exhibit the highest spatial scales of coherence (Figure 4), but O_3 concentrations also exhibit a variance peak at synoptic time scales (though it is smallest at SPL). The short atmospheric lifetime of NH_3 [*Clarisse et al*., 2009], and seasonality in NH_3 fertilizer application and emissions from other agricultural activities [*Aneja et al*., 2003; *Goebes et al*., 2003], is reflected in NH_3 variability being characterized by the synoptic and seasonal modes (Figure 3e), and that seasonality may be in part responsible for the observed seasonality in GR and J_n

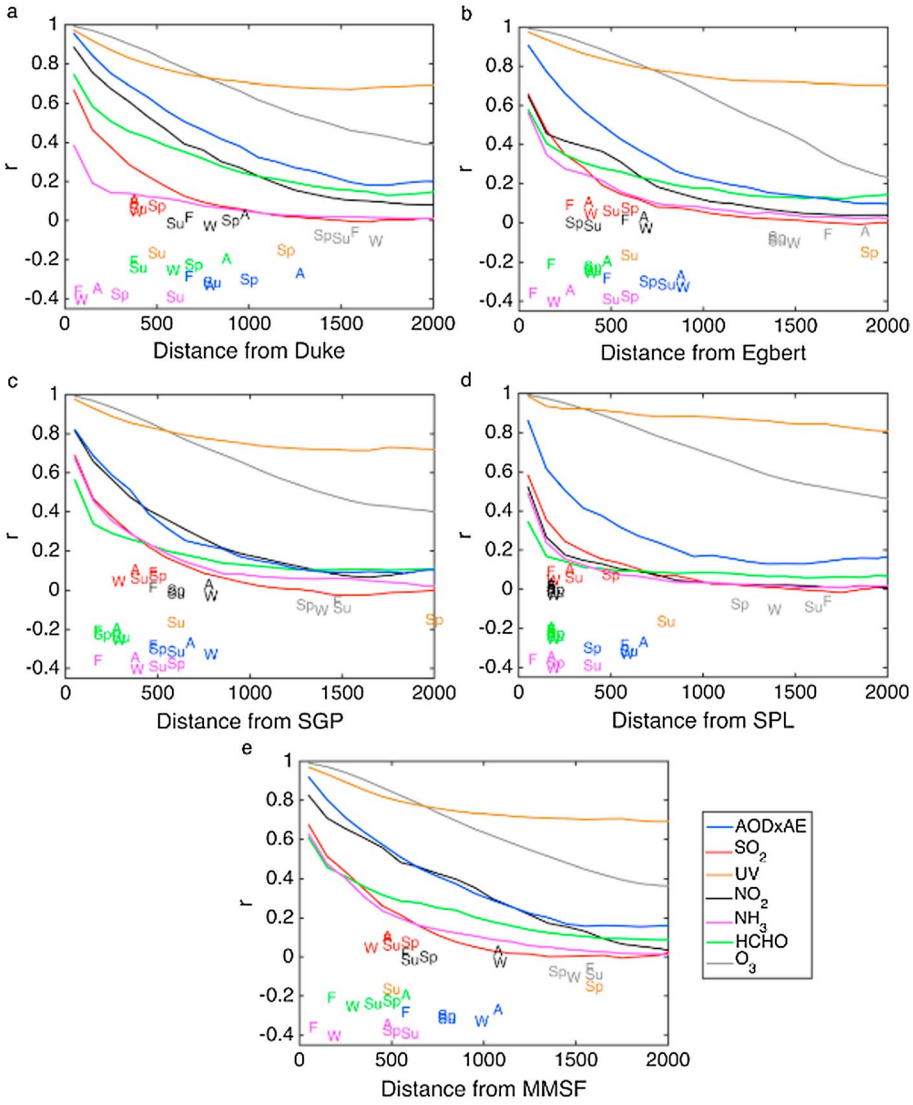

Figure 4. Mean correlation (*r*; binned in 100 km distances) between the time series of each of the predictors for the grid cell containing each PSD measurement site and all other grid cells expressed as a function of separation distance. The mean correlation is thus computed over all azimuth directions. The PSD measurement sites are (a) Duke Forest, (b) Egbert, (c) Southern Great Plains, (d) Storm Peak Laboratory, and (e) Morgan Monroe State Forest. The distance at which $r < 0.3$ by season is indicated by the abscissa location of "*W*" = winter, "Sp" = spring, "Su" = summer, "F" = fall, and "A"= all measurements, with the ordinate location selected solely for visibility. For this analysis the entire satellite measurement period is used. The OMI measurements (SO_2, UV, NO_2, HCHO, and O_3) are from 2004 to 2014, MODIS measurements (AOD × AE) are from 2000 to 2014, and IASI measurements (NH_3) are from 2008 to 2014. Due to low temporal resolution of MODIS LAI (1 in 8 days), LAI × *T* is omitted.

(Figure 2). AOD × AE varies on the annual time scale at all sites and strongly at the synoptic time scale at Egbert and SGP, moderately at SPL and MMSF, and weakly at that time scale at Duke. Many of the variables show a minimum in variance at the ~180 day period, supporting separation of the data into leaf-active and leaf-dormant periods for the regression analysis below.

In accord with previous analyses that suggest that NPF occurs at the regional scale and exhibits temporal autocorrelation, at least for lags of 1 day, the predictors also exhibit relatively large scales of spatial coherence around all PSD measurement sites (Figure 4). Further, consistent with the highest overall probability of NPF and p(1|1) at Duke, all predictors (except NH_3) exhibit comparatively large scales of spatial coherence there (Figure 4a). At all sites the gas-phase concentrations (particularly SO_2, NH_3, and HCHO) exhibit smaller scales of coherence than AOD × AE and UV (Figure 4), potentially indicating that they may play a greater role in

determining the likelihood of NPF events. NH_3 concentrations exhibit greater spatial coherence in the spring and summer, particularly at Egbert, SPL, and MMSF (Figures 4b, 4c, and 4e). This finding is consistent with the higher NH_3 concentrations on event days during the leaf-active period (Table 3) and may contribute to the spring peak in NPF frequency and high observed GR at these sites during these seasons [*Zhang et al*., 2010; *Pryor et al*., 2011]. NO_2 and AOD × AE exhibit similar (and large) spatial scales at MMSF and SGP, implying that anthropogenic primary particle emissions may dominate accumulation mode concentrations and thus the CS (Figures 4c and 4e). HCHO exhibits much larger scales of coherence than SO_2 and NH_3 at Duke due to the large regional isoprene emissions in the southeastern U.S. [*Millet et al*., 2008] as reflected in the high leaf-active HCHO concentrations at this site (Table 3). These local differences in the temporal variability and spatial coherence of the predictor variables may thus offer partial explanations for the site-to-site variations in NPF occurrence and characteristics.

3.3. Association Between Satellite-Based Measurements and NPF Occurrence and Event Characteristics at the Five Sites

Consistent with prior research [*Yu et al*., 2015; *Sullivan and Pryor*, 2016], NPF frequency and characteristics exhibit marked seasonality (Figure 2) as does the dependence on the satellite-based predictors. For example, NPF is more frequent when LAI × T is lower during leaf-active and higher during leaf-dormant measurement days (Table 3), due to the higher frequency of events during spring and fall (Figure 2). The dominant difference in the satellite-based predictors on NPF event versus nonevent days during leaf-dormant periods is insolation receipt (UV) and the resultant production of atmospheric oxidants, while the differences in the predictors during the leaf-active periods are more complex. Consistent with the expectation that a higher CS will tend to suppress NPF, nonevent days are characterized by higher AOD × AE at all sites during the leaf-active period (Table 3). However, this is not the case during the leaf-dormant season when particle loading is generally lower (except SPL; Table 3) [*Sullivan et al*., 2015], indicating that another parameter(s) (e.g., availability and/or oxidation of NPF precursors) may limit NPF during leaf-dormant periods. Despite the large uncertainty in NH_3 retrievals, consistent with a priori expectations of ternary nucleation, higher concentrations are observed on NPF event days at the two sites located near high NH_3 emissions, SGP (significant difference leaf-active; $\alpha = 0.1$) and MMSFb (higher p25 and p75 leaf-active; significant difference leaf-dormant) [*Goebes et al*., 2003; *U.S. Environmental Protection Agency*, 2011]. Higher HCHO concentrations are observed on leaf-active nonevent days at all sites (not significant at Egbert, p-value = 0.11), supportive of the postulate that isoprene (likely the major source of remotely sensed HCHO) tends to quench available oxidants, reduce H_2SO_4 production, and suppress NPF. There was no significant difference in HCHO on event and nonevent days during leaf-dormant periods, supporting the assertion that satellite-based measurements of HCHO are primarily indicative of BVOC emissions. Significantly higher O_3 concentrations are observed on NPF days during the leaf-active season (though not at Duke or Egbert) possibly indicating that high non-isoprene VOC concentrations are associated with both high O_3 production and an increased likelihood of NPF.

Regression trees constructed to "predict" event occurrence illustrate the importance of predictor interactions (Figure 5). For example, in the MMSFa data set the overall NPF frequency is 20%, but increases to 30% when $UV > 26\,mW\,m^{-2}\,nm^{-1}$ and to 39% when $UV > 26\,mW\,m^{-2}\,nm^{-1}$ and $LAI \times T < 580\,m^2\,m^{-2}\,K$ (Figure 5e). At Duke, the first node is LAI × T ≈ median (Table 3), with an increase in NPF frequency from 26% to 64% conditional on AOD × AE < 0.23 if $LAI \times T > 670\,m^2\,m^{-2}\,K$ and from 54% to 68% when $UV > 28\,mW\,m^{-2}\,nm^{-1}$ if $LAI \times T < 670\,m^2\,m^{-2}\,K$ (cf. 41% for all days; Figure 5a). This is consistent with the event versus nonevent day conditions described above, where AOD × AE and UV are important discriminators between event and nonevent days during leaf-active and leaf-dormant period, respectively (Table 3). AOD × AE is the first or second level node variable at all sites except MMSFa, and in all cases low AOD × AE is associated with increased probability of NPF (Figures 5a–5d and 5f). LAI × T and UV are also important discriminators of event and nonevent days, each being the first or second level node at three of the six sites. Higher UV is typically associated with increased probability of NPF, while the relationship with LAI × T is less clear because, as discussed above, NPF frequency is highest at moderate LAI × T (i.e., in spring and fall). O_3 is the first node and NO_2 is a second level node at SPL, where higher O_3 and lower NO_2 favors NPF, indicating that the presence of high concentrations of stabilizing organics coinciding with low anthropogenic emissions is favorable for NPF (Figure 5d). MMSF is in a location of high BVOC and NH_3 emissions. Accordingly, NH_3 and HCHO are the second level nodes with higher NH_3 and lower HCHO being associated with increased probability of NPF at MMSFb

Table 3. Median [p25-p75] Conditions From Satellite-Based Measurements on Event and Nonevent Days (*n*) During Leaf-Active and Leaf-Dormant Measurements[a]

		Duke		Egbert		Southern Great Plains	
		Event	Nonevent	Event	Nonevent	Event	Nonevent
Leaf-active	n	64	144	55	122	83	655
	AODxAE	**0.47 [0.14–0.73]**	**0.68 [0.43–0.92]**	**0.26 [0.14–0.37]**	**0.32 [0.22–0.55]**	**0.10 [0.03–0.18]**	**0.15 [0.07–0.25]**
	SO_2 (DU)	0.06 [−0.19–0.02]	−0.05 [−0.21–0.06]	**0.08 [−0.06–0.21]**	**−0.05 [−0.22–0.11]**	−0.02 [−0.20–0.14]	−0.01 [−0.15–0.11]
	UV ($mW\,m^{-2}\,nm^{-1}$)	96 [74–106]	97 [86–105]	**82 [69–90]**	**75 [50–91]**	**84 [67–109]**	**96 [75–111]**
	NO_2 ($\times10^{15}$ molec. cm^{-2})	2.91 [2.13–3.94]	2.64 [2.22–3.14]	2.53 [1.37–3.93]	2.11 [1.33–3.53]	1.72 [1.30–2.11]	1.65 [1.28–2.05]
	NH_3 ($\times10^{15}$ molec. cm^{-2})	N/A	N/A	2.32 [2.32–2.32]	2.32 [2.32–2.32]	**8.06 [4.21–13.4]**	**6.41 [2.68–10.9]**
	HCHO ($\times10^{16}$ molec. cm^{-2})	**1.53 [1.17–1.87]**	**1.79 [1.49–2.25]**	1.00 [0.82–1.26]	1.11 [0.86–1.35]	**1.05 [0.89–1.32]**	**1.16 [0.94–1.43]**
	LAIxT ($m^2\,m^{-2}$ K)	**1053 [665–1197]**	**1140 [1035–1207]**	594 [401–769]	682 [394–784]	**322 [248–405]**	**351 [268–423]**
	O_3 (DU)	315 [301–328]	311 [303–326]	311 [294–336]	316 [294–337]	**313 [298–331]**	**305 [295–322]**
Leaf-dormant	n	108	99	48	130	79	657
	AODxAE	0.11 [0.06–0.23]	0.10 [0.05–0.19]	0.20 [0.10–0.27]	0.20 [0.14–0.25]	0.07 [0.03–0.16]	0.09 [0.04–0.17]
	SO_2 (DU)	**0.08 [−0.10–0.26]**	**−0.03 [−0.23–0.19]**	**−0.09 [−0.53–0.12]**	**0.02 [−0.36–0.29]**	0.02 [−0.21–0.21]	−0.02 [−0.24–0.17]
	UV ($mW\,m^{-2}\,nm^{-1}$)	**47 [36–64]**	**29 [21–42]**	**19 [10–35]**	**12 [7–22]**	**37 [31–57]**	**35 [25–51]**
	NO_2 ($\times10^{15}$ molec. cm^{-2})	5.23 [4.15–6.71]	5.46 [4.32–6.85]	**3.36 [1.49–5.25]**	**3.95 [2.25–6.65]**	2.53 [2.01–3.31]	2.49 [1.93–3.14]
	NH_3 ($\times10^{15}$ molec. cm^{-2})	N/A	N/A	2.32 [1.85–2.32]	2.32 [−0.31–2.41]	3.89 [1.92–9.00]	4.17 [1.49–8.71]
	HCHO ($\times10^{16}$ molec. cm^{-2})	0.96 [0.82–1.12]	0.96 [0.83–1.12]	0.84 [0.64–1.17]	0.84 [0.63–0.97]	1.12 [0.92–1.26]	1.04 [0.90–1.20]
	LAIxT ($m^2\,m^{-2}$ K)	**288 [223–312]**	**245 [236–298]**	23 [13–165]	19 [15–64]	**149 [139–170]**	**141 [110–160]**
	O_3 (DU)	299 [280–324]	299 [264–326]	319 [289–365]	336 [300–366]	**281 [266–302]**	**294 [274–318]**

[a]Values in bold indicate rejection of the null hypothesis that the samples are from the same population ($\alpha = 0.1$, Wilcoxon rank sum test). Note that NH_3 measurements are available beginning in 2008 and are thus available for only a portion of the PSD measurements at Egbert and MMSFa (no coincident measurements at Duke), and days without coincident measurements (e.g., all of 2007) and are filled with a mean NH_3 value. Thus, the distributions of NH_3 on event versus nonevent days can be significantly different, despite the median values being strongly driven by missing data.

(NH_3 measurements are not available for most of MMSFa), again emphasizing the role of a stabilizing base (such as NH_3) in promoting NPF and supporting the postulate that high isoprene emissions can suppress NPF (Figure 5f).

In general, multiple linear regression models constructed by using the satellite-derived variables as predictors and NPF characteristics as predictands explain more of the variability in GR, J_n, SP, and $N_{n-100\,nm}$ at each site than a random model with equal sample size and number of predictors, indicating that the satellite-based predictors exhibit some explanatory skill in characterizing NPF events over North America (Figure 6). The multiple linear regression models exhibit higher explanatory power for J_n, SP, and $N_{n-100\,nm}$ than for GR, indicating that the proxy variables are better able to capture the intensity of NPF events than the growth rates (Figure 6), potentially because the species that participate in nucleation and subsequent growth may differ [*Kulmala et al.*, 2004]. Further, although there is some site-to-site consistency in terms of which predictor variables have significant coefficients in the models, the absolute form of the regression models is variable from site-to-site, and generally, the r^2 of the regression models is higher in the leaf-dormant periods (Figures 6 and 7).

At all sites UV has a positive β_i for regression equations of J_n (and $N_{n-100\,nm}$) particularly in leaf-dormant periods. Consistent with higher UV on leaf-dormant NPF event days (Table 3), this indicates that UV not

Table 3. (continued).

		Storm Peak Laboratory		Morgan Monroe State Forest a		Morgan Monroe State Forest b	
		Event	Nonevent	Event	Nonevent	Event	Nonevent
Leaf-active	n	68	219	66	183	54	152
	AODxAE	**0.14 [0.09–0.22]**	**0.21 [0.13–0.29]**	**0.19 [0.11–0.35]**	**0.32 [0.16–0.59]**	**0.05 [−0.02–0.26]**	**0.22 [0.08–0.35]**
	SO_2 (DU)	0.02 [−0.10–0.18]	0.09 [−0.05–0.20]	**−0.03 [−0.19–0.12]**	**0.04 [−0.07–0.18]**	0.01 [−0.16–0.16]	0.00 [−0.13–0.09]
	UV ($mW\,m^2\,nm^{-1}$)	**129 [124–134]**	**125 [115–134]**	68 [54–98]	72 [44–97]	**91 [76–100]**	**86 [68–99]**
	NO_2 ($\times10^{15}$ mol cm^{-2})	0.75 [0.52–0.95]	0.69 [0.48–0.87]	**3.93 [3.10–5.13]**	**3.41 [2.21–4.73]**	**2.54 [1.92–3.30]**	**2.11 [1.69–2.68]**
	NH_3 ($\times10^{15}$ mol cm^{-2})	1.20 [0.65–2.30]	1.63 [0.45–3.21]	0.18 [0.18–0.18]	0.18 [0.18–0.18]	6.16 [3.17–10.9]	6.37 [2.82–10.4]
	HCHO ($\times10^{16}$ mol cm^{-2})	**1.15 [0.84–1.64]**	**1.40 [1.08–1.82]**	**1.06 [0.91–1.20]**	**1.25 [0.96–1.56]**	**1.19 [0.98–1.72]**	**1.51 [1.20–1.80]**
	LAIxT ($m^2\,m^{-2}$ K)	332 [271–415]	359 [277–390]	**188 [130–462]**	**512 [182–855]**	**546 [486–673]**	**658 [508–689]**
	O_3 (DU)	**314 [299–323]**	**291 [283–297]**	**315 [295–337]**	**305 [284–325]**	**311 [302–329]**	**304 [296–314]**
Leaf-dormant	n	86	197	29	203	60	146
	AODxAE	0.23 [0.21–0.26]	0.23 [0.17–0.27]	0.14 [0.09–0.20]	0.14 [0.11–0.17]	0.08 [0.02–0.15]	0.10 [0.05–0.17]
	SO_2 (DU)	**−0.03 [−0.27–0.18]**	**−0.15 [−0.38–0.05]**	−0.13 [−0.43–0.23]	−0.13 [−0.46–0.18]	**−0.20 [−0.40–0.07]**	**−0.05 [−0.36–0.20]**
	UV ($mW\,m^2\,nm^{-1}$)	**78 [47–104]**	**50 [28–74]**	**29 [26–41]**	**15 [10–23]**	**43 [27–63]**	**27 [18–38]**
	NO_2 ($\times10^{15}$ mol cm^{-2})	**0.39 [0.28–0.57]**	**0.56 [0.35–1.00]**	5.01 [3.32–7.04]	4.79 [3.06–5.89]	3.86 [3.14–5.39]	4.17 [3.25–5.31]
	NH_3 ($\times10^{15}$ mol cm^{-2})	1.23 [−0.06–3.35]	1.62 [−0.05–3.91]	0.18 [0.18–0.76]	0.18 [−2.92–0.70]	**4.26 [0.87–9.77]**	**1.75 [−0.51–6.68]**
	HCHO ($\times10^{16}$ mol cm^{-2})	1.03 [0.93–1.31]	1.04 [0.91–1.27]	0.88 [0.70–1.11]	0.86 [0.72–1.01]	1.06 [0.91–1.17]	1.00 [0.84–1.18]
	LAIxT ($m^2\,m^{-2}$ K)	49 [36–97]	61 [40–118]	**86 [62–95]**	**74 [54–85]**	**163 [115–287]**	**100 [83–141]**
	O_3 (DU)	**308 [287–336]**	**295 [275–332]**	315 [297–336]	310 [286–339]	292 [275–305]	296 [278–314]

only controls whether NPF occurs but also the intensity (J_n). Conversely, increased UV is associated with decreased SP, consistent with higher formation rates increasing particle loss through self-coagulation and reducing survival probability of individual particles. In general, β_i for AOD × AE are negative in equations for J_n and $N_{n-100\,nm}$, particularly during the leaf-active period. AOD × AE has a significant positive β_i for SP during leaf-active events at MMSFa, but a negative β_i for SP at SGP and MMSFb, indicating the controls on survival probabilities may vary both in space and time. Higher AOD × AE is expected to reduce SP by increasing coagulation loss, but as discussed above AOD × AE appears to be driven by anthropogenic emissions at MMSF and SGP (Figures 4c and 4e), and thus may also indicate the presence of high precursor concentrations. Increased precursor concentrations can increase GR (e.g., AOD × AE exhibits positive β_i for GR at MMSFa) and therefore increase SP, which may explain the positive association between AOD × AE and SP at MMSFa and the lack of AOD × AE dependence in the MMSFa regression tree (Figure 5). The sign and significance of the β_i weights on SO_2, NO_2, NH_3, HCHO, and O_3 are highly variable by site and leaf activity (Tables S1 and S2 in the supporting information), which may reflect differential NPF mechanisms in space and time, and thus explain the site-to-site differences in NPF frequency and characteristics. For example, SO_2 has a positive, significant β_i for $N_{n-100\,nm}$ at SPL (leaf-active) and Duke (leaf-dormant), but generally negative β_i elsewhere; NO_2 has a negative, significant β_i for $N_{n-100\,nm}$ at Duke (leaf-active), but positive, significant β_i at Egbert (leaf-active), and variable sign elsewhere, and NH_3, HCHO, and O_3 are generally split between positive and negative β_i across the sites for all NPF metrics. As each of the predictor variables are significant at at least one site, we retain them

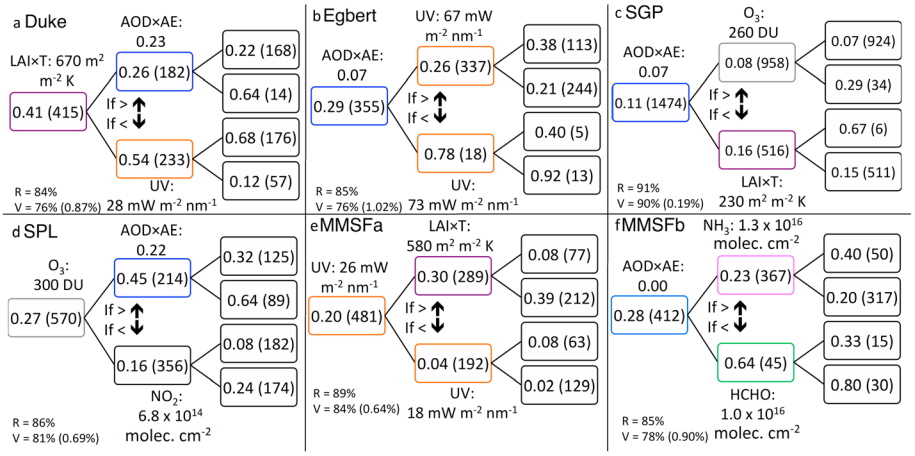

Figure 5. Regression trees for predicting NPF event occurrence at (a) Duke, (b) Egbert, (c) SGP, (d) SPL, (e) MMSFa, and (f) MMSFb. The branches upward and downward are for all days above and below the variable threshold given at the node, respectively. The colors of the boxes correspond to the variable colors from Figure 4, with the addition that purple is used for LAI × T. The probability of an event is given at each node, and the sample size is given in the parenthesis (note the far left nodes are the entire data set at each site). Also given is the resubstitution accuracy (*R*) and mean (standard deviation) cross-validation accuracy (*V*) after withholding 20% of measurements days, over 1000 iterations. The trees are built using a maximum of 10 nodes and minimum leaf size of 5, but have been truncated here for legibility.

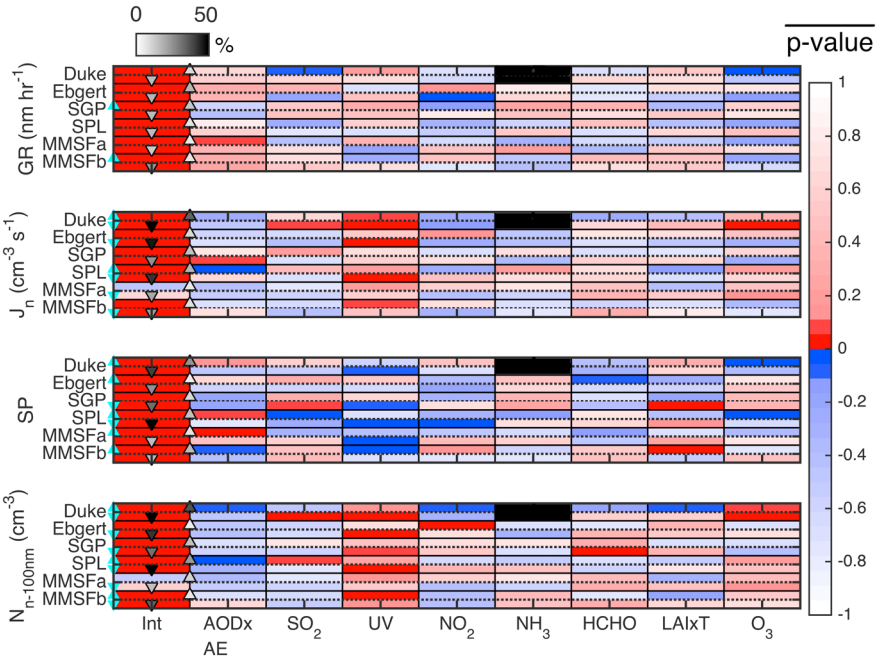

Figure 6. *P*-value of the regression coefficients on each of the predictor variables averaged across the five training/validation cross-validation data sub sets and across the 1000 iterations of subsampling, used in the multiple linear regression (equation (7)) to predict event metrics (growth rate, formation rate, and survival probability) and daily mean ultrafine particle ($D_p < 100$ nm; $N_{n\ -\ 100\ nm}$, where n is the instrument minimum D_p detection limit) concentrations. Satellite-based measurements of NH_3 are not available prior to 2008, and thus are not available for Duke (black fill) and portions of the Egbert and MMSFa PSD measurement days. For each site, the first and second rows are leaf-active and leaf-dormant periods, respectively. The red and blue indicate a positive and negative coefficient weighting, respectively, with the opacity indicating the significance (mean p-value) of the weighing. A cyan triangle on the ordinate indicates a significant *p*-value ($\alpha = 0.1$) for the regression model trained on the complete data set, and full model coefficients are given in Tables S1 and S2. The shading of the second (black) triangles (upward = leaf-active, downward = leaf-dormant; abscissa offset solely for visibility) is the percentage of data subsets that show significantly ($\alpha = 0.1$) higher r^2 in predicting the NPF characteristics than expected by random chance (*F*-statistic for the sample size and number of predictors).

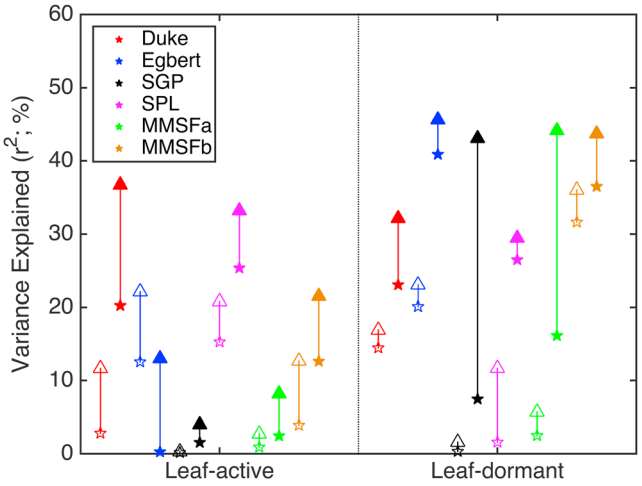

Figure 7. Variance of daily mean UFP particle concentration ($N_{n-100\,nm}$) explained (r^2) for each data set using a simple model (stars; where in the predictors are: AOD × AE, SO_2, and UV [*Kulmala et al.*, 2011; *Crippa et al.*, 2013; *Sundström et al.*, 2015]) and the full model developed herein (triangles; wherein the predictors are AOD × AE, SO_2, UV, NO_2, NH_3, HCHO, LAI × *T*, and O_3), during all measurement days (open symbols) and during only event days (filled symbols).

all, but site-to-site variability in which predictor variable(s) are most important suggests a key challenge in building a generalizable model.

3.4. Impact of Subsampling on Stability of Analyses

The regression trees described in section 3.3 were built by using all measurement days but were also built after withholding 20% of the data, for cross-validation analysis. The full model regression trees had resubstitution accuracies of 85–91%. When the testing data are withheld from the training models, the cross-validation accuracies averaged 76–90% (mean standard deviation of accuracies of ~<1%) across 1000 iterations of cross validation (Figure 5). Thus, subsampling only moderately reduces the models' accuracy, and the results are relatively stable independent of the specific subsampled days.

Running the multiple linear regressions as Monte Carlo experiments, and iterating the cross-validation analysis, provides insights into the models' stability. Without cross validation, multiple linear regression models explain a significant amount of the observed variability in the NPF characteristics (see cyan triangles in Figure 6). However, when the cross validation is performed, it is not uncommon that the training data sets show poor performance in predicting characteristics (particularly for GR; Figure 6) when some of the validation data were withheld from training the models. This implies that the models are not generalizable. Conversely, the models for J_n and $N_{n-100\,nm}$ at Duke, Egbert (leaf-dormant), and SPL (leaf-dormant), and SP at Duke (leaf-dormant) and SPL (leaf-dormant), seem robust, independent of the data set subsampling, indicating that the precise measurement dates may not significantly impact the inferences drawn herein.

3.5. Improved Satellite-Based Proxy for Ultrafine Particle Concentrations

Both the simple proxy model, where UFP total number concentrations ($N_{n-100\,nm}$) are predicted by using only AOD × AE, SO_2, and UV as predictors, and the full models, which include additional predictor variables: NO_2, NH_3, HCHO, LAI × *T*, and O_3, have much higher explanatory skill when only NPF event days are considered (Figure 7). The full model improves explanatory skill over the simpler model at all sites during both leaf-active and leaf-dormant periods (Figure 7) and is associated with variance explanation (r^2) on NPF event days of 29–46% during the leaf-dormant period and 4–37% during leaf-active periods (Figure 7). The increase in variance explanation with the addition of extra predictors is particularly large at Duke, SGP, and MMSF. These are locations influenced by high organic emissions (quantified using HCHO, LAI × *T*, and O_3), anthropogenic emissions (quantified using NO_2), and ternary nucleation precursor emissions (e.g., NH_3) (Figure 1) [*Goebes et al.*, 2003] and may be indicative of an enhanced role of these species in dictating UFP concentrations at these sites.

4. Discussion and Conclusions

We examine the frequency, persistence, and characteristics of NPF events at five locations across North America and employ statistical analysis of satellite-based measurements of atmospheric composition to explain spatial similarities and variability in NPF frequency, autocorrelation, formation rates, growth rates, survival probabilities, and daily mean ultrafine particle concentrations. Despite large geographic separation, and vastly different local land use and point source pollution emissions between the sites, NPF is observed at all sites with peak frequencies in spring and fall and exhibits positive 1 day autocorrelation. Accordingly, the

temporal modes of variability and spatial scales of coherence of the remotely sensed variables thought to control NPF also show considerable site-to-site consistency and are typically coherent on larger scales at sites with larger 1 day autocorrelation in NPF occurrence. There is also broad agreement in terms of the conditions associated with NPF events between the sites: NPF is more frequent during moderate LAI × *T*, low AOD × AE, and low HCHO during leaf-active periods, and high UV in leaf-dormant periods. The spatial consistencies in the primary drivers of NPF may explain why simplified nucleation schemes can be used with some skill in global models to characterize the impact of NPF on particle size distributions and CCN concentrations [*Spracklen et al*., 2008b]. However, event characteristics (J_n, GR, SP, and $N_{n-100\,nm}$) exhibit greater site-to-site variability in terms of their dependence on the remote-sensing predictors. Site-to-site variability in NPF characteristics and the corresponding variability in satellite-based measurements of the drivers of GR, J_n, SP, and $N_{n-100\,nm}$ may explain the spatial variability in the performance of simplified NPF schemes [*Lee et al*., 2013]. Generalized schemes with a single NPF mechanism and set of coefficients may not be able to capture the variability in precise nucleation mechanisms [*Yu et al*., 2015] and/or importance of specific precursor species at the different sites. Proxy models of total UFP concentrations that expand the suite of remote sensing predictors exhibit improved variance explanation relative to simpler models that have been previously proposed [*Kulmala et al*., 2011; *Crippa et al*., 2013; *Sundström et al*., 2015] (Figure 7). However, the model coefficients and hence the magnitude and even sign of the dependencies of $N_{n-100\,nm}$ on the suite of predictors considered (AOD × AE, SO_2, UV, NO_2, NH_3, HCHO, LAI × T, O_3) (Figure 6) imply great challenges to generating a single generalizable proxy. We recommend future NPF schemes try to reproduce the (spatially variable) relationships between NPF and its drivers presented here, and connect existing theory with ground and satellite observations to evaluate new model treatments of NPF before accepting them, and ultimately improve understanding of regional to global scale impacts of NPF on climate.

Future research is necessary to examine how satellite-based measurement error impacts the explanatory skill (and variability) of the proxies and to further examine the feasibility of developing global proxies for NPF occurrence and characteristics. Further, it may be appropriate to develop the proxies using nonlinear techniques or additional (or compound) variables. Given the high uncertainty in direct satellite measurements, the proxies may benefit from use of variables from reanalysis products such as Modern Era Retrospective Analysis for Research and Applications version 2 [*Bosilovich et al*., 2016] or output from satellite-constrained chemical transport models (CTM) since a number of CTM now exhibit skill for many of the predictor variables used here [*Westervelt et al*., 2013]. If such a proxy could be found it may provide computationally efficient first-order estimates of the impact of NPF on particle size distributions, CCN concentrations, and ultimately the potential impact of NPF on climate.

Acknowledgments

Funding was supplied by the NASA Earth and Space Science Fellowship Program Grant "14-EARTH14F-0207" (R.C.S.), L'Oréal-UNESCO UK and Ireland Fellowship For Women In Science (P.C.), Belgian F.R.S-FNRS (L.C.), Boursier FRIA (S.W.), NSF (1102309 and 1517365), and NASA (NNX16AG31G) (S.C.P.). We also gratefully acknowledge computational resources provided by the Lilly Endowment, Inc. to the Indiana University Pervasive Technology Institute and the Indiana METACyt Initiative. The MODIS and OMI satellite data were collected under the NASA Earth-Sun System Division and Science Mission Directorate, SGP PSD data were acquired under the DoE Atmospheric Radiation Measurement Program. SPL is a permittee of the Medicine-Bow Routt National Forests, and we acknowledge assistance from Randolph Borys and Ian McCubbin. The views expressed in this article are those of the author(s) and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency. Data are available from: MODIS and OMI satellite data (NASA, http://reverb.echo.nasa.gov/reverb), IASA NH_3 (L. Clarisse, lclariss@ulb.ac.be), PSD from Duke (J. T. Walker, Walker.Johnt@epa.gov), Egbert (W. R. Leaitch, richard.leaitch@canada.ca), SGP (U.S. DoE, http://www.archive.arm.gov/), SPL (A. G. Hallar, gannet.hallar@dri.edu), and MMSF (S. C. Pryor, sp2279@cornell.edu).

References

Ackerman, T. P., and G. M. Stokes (2003), The atmospheric radiation measurement program, *Phys. Today*, *56*(PNNL-SA-37894), 38–44.

Almeida, J., S. Schobesberger, A. Kürten, I. K. Ortega, O. Kupiainen-Määttä, A. P. Praplan, A. Adamov, A. Amorim, F. Bianchi, and M. Breitenlechner (2013), Molecular understanding of sulphuric acid-amine particle nucleation in the atmosphere, *Nature*, *502*(7471), 359–363.

Anderson, T. L., R. J. Charlson, D. M. Winker, J. A. Ogren, and K. Holmén (2003), Mesoscale variations of tropospheric aerosols, *J. Atmos. Sci.*, *60*(1), 119–136.

Aneja, V. P., D. R. Nelson, P. A. Roelle, J. T. Walker, and W. Battye (2003), Agricultural ammonia emissions and ammonium concentrations associated with aerosols and precipitation in the southeast United States, *J. Geophys. Res.*, *108*, 4152, doi:10.1029/2002JD002271.

Ångström, A. (1964), The parameters of atmospheric turbidity, *Tellus*, *16*(1), 64–75.

Bosilovich, M. G., R. Lucchesi, and M. Suarez (2016), MERRA-2: File specification. GMAO Office Note No. 9 (Version 1.1), 73 pp. [Available at http://gmao.gsfc.nasa.gov/pubs/office_notes.]

Boy, M., B. Bonn, J. Kazil, N. Lovejoy, A. Turnipseed, J. Greenberg, T. Karl, L. Mauldin, E. Kusciuch, and J. Smith (2007), Relevance of several nucleation theories in different environments, in *Nucleation and Atmospheric Aerosols*, pp. 87–91, Springer, Dordrecht, Netherlands.

Brinksma, E. J., K. F. Boersma, P. F. Levelt, and R. D. McPeters (2003), OMI validation requirements document, version 1, *Rep. RS-OMIE-KNMI-345*, 66 pp., [Available at http//projects.knmi.nl/omi/documents/validation/RS-OMIE-KNMI-345_i1.pdf.]

Carslaw, K. S., L. A. Lee, C. L. Reddington, K. J. Pringle, A. Rap, P. M. Forster, G. W. Mann, D. V. Spracklen, M. T. Woodhouse, and L. A. Regayre (2013), Large contribution of natural aerosols to uncertainty in indirect forcing, *Nature*, *503*(7474), 67–71.

Chance, K. (2002), OMI algorithm theoretical basis document, volume IV: OMI trace gas algorithms. [Available at http://eospso.gsfc.nasa.gov/sites/default/files/atbd/ATBD-OMI-04.pdf.]

Chance, K., P. I. Palmer, R. J. D. Spurr, R. V. Martin, T. P. Kurosu, and D. J. Jacob (2000), Satellite observations of formaldehyde over North America from GOME, *Geophys. Res. Lett.*, *27*(21), 3461–3464, doi:10.1029/2000GL011857.

Clarisse, L., C. Clerbaux, F. Dentener, D. Hurtmans, and P.-F. Coheur (2009), Global ammonia distribution derived from infrared satellite observations, *Nat. Geosci.*, *2*(7), 479–483.

Crippa, P., and S. C. Pryor (2013), Spatial and temporal scales of new particle formation events in eastern North America, *Atmos. Environ.*, *75*, 257–264, doi:10.1016/j.atmosenv.2013.04.051.

Crippa, P., D. Spracklen, and S. C. Pryor (2013), Satellite-derived estimates of ultrafine particle concentrations over eastern North America, *J. Geophys. Res. Atmos.*, *118*, 9968–9981, doi:10.1002/jgrd.50707.

Fang, H., S. Wei, and S. Liang (2012), Validation of MODIS and CYCLOPES LAI products using global field measurement data, *Remote Sens. Environ.*, *119*, 43–54.

Fioletov, V. E., C. A. McLinden, N. Krotkov, M. D. Moran, and K. Yang (2011), Estimation of SO_2 emissions using OMI retrievals, *Geophys. Res. Lett.*, *38*, L21811, doi:10.1029/2011GL049402.

Goebes, M. D., R. Strader, and C. Davidson (2003), An ammonia emission inventory for fertilizer application in the United States, *Atmos. Environ.*, *37*(18), 2539–2550.

Guenther, A. B., P. R. Zimmerman, P. C. Harley, R. K. Monson, and R. Fall (1993), Isoprene and monoterpene emission rate variability: Model evaluations and sensitivity analyses, *J. Geophys. Res.*, *98*(D7), 12,609–12,617, doi:10.1029/93JD00527.

Hallar, A. G., D. H. Lowenthal, G. Chirokova, R. D. Borys, and C. Wiedinmyer (2011), Persistent daily new particle formation at a mountain-top location, *Atmos. Environ.*, *45*(24), 4111–4115.

Hallar, A. G., R. Petersen, I. B. McCubbin, D. Lowenthal, S. Lee, E. Andrews, and F. Yu (2016), Climatology of new particle formation and corresponding precursors at Storm Peak Laboratory, *Aerosol Air Qual. Res.*, *16*(3), 816–826.

Helmig, D. (1997), Ozone removal techniques in the sampling of atmospheric volatile organic trace gases, *Atmos. Environ.*, *31*(21), 3635–3651.

Henze, D. K., and J. H. Seinfeld (2006), Global secondary organic aerosol from isoprene oxidation, *Geophys. Res. Lett.*, *33*, L09812, doi:10.1029/2006GL025976.

Hussein, T., H. Junninen, P. Tunved, A. Kristensson, M. Dal Maso, I. Riipinen, P. P. Aalto, H.-C. Hansson, E. Swietlicki, and M. Kulmala (2009), Time span and spatial scale of regional new particle formation events over Finland and Southern Sweden, *Atmos. Chem. Phys.*, *9*(14), 4699–4716.

Hyvönen, S., H. Junninen, L. Laakso, M. D. Maso, T. Grönholm, B. Bonn, P. Keronen, P. Aalto, V. Hiltunen, and T. Pohja (2005), A look at aerosol formation using data mining techniques, *Atmos. Chem. Phys.*, *5*(12), 3345–3356.

Jeong, C.-H., G. J. Evans, M. L. McGuire, R.-W. Chang, J. P. D. Abbatt, K. Zeromskiene, M. Mozurkewich, S.-M. Li, and W. R. Leaitch (2010), Particle formation and growth at five rural and urban sites, *Atmos. Chem. Phys.*, *10*(16), 7979–7995.

Jokinen, T., T. Berndt, R. Makkonen, V.-M. Kerminen, H. Junninen, P. Paasonen, F. Stratmann, H. Herrmann, A. B. Guenther, and D. R. Worsnop (2015), Production of extremely low volatile organic compounds from biogenic emissions: Measured yields and atmospheric implications, *Proc. Natl. Acad. Sci. U.S.A.*, *112*(23), 7123–7128.

Kiendler-Scharr, A., J. Wildt, M. Dal Maso, T. Hohaus, E. Kleist, T. F. Mentel, R. Tillmann, R. Uerlings, U. Schurr, and A. Wahner (2009), New particle formation in forests inhibited by isoprene emissions, *Nature*, *461*(7262), 381–384.

Kirkby, J., J. Curtius, J. Almeida, E. Dunne, J. Duplissy, S. Ehrhart, A. Franchin, S. Gagné, L. Ickes, and A. Kürten (2011), Role of sulphuric acid, ammonia and galactic cosmic rays in atmospheric aerosol nucleation, *Nature*, *476*(7361), 429–433.

Knol, A. B., J. J. de Hartog, H. Boogaard, P. Slottje, J. P. van der Sluijs, E. Lebret, F. R. Cassee, J. A. Wardekker, J. G. Ayres, and P. J. Borm (2009), Expert elicitation on ultrafine particles: Likelihood of health effects and causal pathways, *Part. Fibre Toxicol.*, *6*(1), 19, doi:10.1186/1743-8977-6-19.

Krotkov, N. A., S. A. Carn, A. J. Krueger, P. K. Bhartia, and K. Yang (2006), Band residual difference algorithm for retrieval of SO_2 from the Aura Ozone Monitoring Instrument (OMI), *IEEE Trans. Geosci. Remote Sens.*, *44*(5), 1259–1266.

Krotkov, N. A., B. McClure, R. R. Dickerson, S. A. Carn, C. Li, P. K. Bhartia, K. Yang, A. J. Krueger, Z. Li, and P. F. Levelt (2008), Validation of SO_2 retrievals from the Ozone Monitoring Instrument over NE China, *J. Geophys. Res.*, *113*, D16S40, doi:10.1029/2007JD008818.

Kuang, C., P. H. McMurry, A. V. McCormick, and F. L. Eisele (2008), Dependence of nucleation rates on sulfuric acid vapor concentration in diverse atmospheric locations, *J. Geophys. Res.*, *113*, D10209, doi:10.1029/2007JD009253.

Kulmala, M., L. Pirjola, and J. M. Mäkelä (2000), Stable sulphate clusters as a source of new atmospheric particles, *Nature*, *404*(6773), 66–69.

Kulmala, M., H. Vehkamäki, T. Petäjä, M. Dal Maso, A. Lauri, V.-M. Kerminen, W. Birmili, and P. H. McMurry (2004), Formation and growth rates of ultrafine atmospheric particles: A review of observations, *J. Aerosol Sci.*, *35*(2), 143–176.

Kulmala, M., A. Arola, T. Nieminen, L. Riuttanen, L. Sogacheva, G. de Leeuw, V.-M. Kerminen, and K. E. J. Lehtinen (2011), The first estimates of global nucleation mode aerosol concentrations based on satellite measurements, *Atmos. Chem. Phys.*, *11*(21), 10,791–10,801.

Kulmala, M., J. Kontkanen, H. Junninen, K. Lehtipalo, H. E. Manninen, T. Nieminen, T. Petäjä, M. Sipilä, S. Schobesberger, and P. Rantala (2013), Direct observations of atmospheric aerosol nucleation, *Science*, *339*(6122), 943–946.

Lamsal, L. N., R. V. Martin, A. Van Donkelaar, M. Steinbacher, E. A. Celarier, E. Bucsela, E. J. Dunlea, and J. P. Pinto (2008), Ground-level nitrogen dioxide concentrations inferred from the satellite-borne Ozone Monitoring Instrument, *J. Geophys. Res.*, *113*, D16308, doi:10.1029/2007JD009235.

Land Processes Distributed Active Archive Center (LP DAAC) (2014), Land Cover Type Yearly L3 Global 0.05Deg CMG, V051. NASA EOSDIS LP DAAC, USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota (https://lpdaac.usgs.gov), accessed 11/13/2015, at 10.5067/MODIS/MCD12C1.006,

Lee, L. A., K. J. Pringle, C. L. Reddington, G. W. Mann, P. Stier, D. V. Spracklen, J. R. Pierce, and K. S. Carslaw (2013), The magnitude and causes of uncertainty in global model simulations of cloud condensation nuclei, *Atmos. Chem. Phys.*, *13*(17), 8879–8914.

Levy, R. C., L. A. Remer, R. G. Kleidman, S. Mattoo, C. Ichoku, R. Kahn, and T. F. Eck (2010), Global evaluation of the Collection 5 MODIS dark-target aerosol products over land, *Atmos. Chem. Phys.*, *10*(21), 10,399–10,420.

Levy, R. C., S. Mattoo, L. A. Munchak, L. A. Remer, A. M. Sayer, F. Patadia, and N. C. Hsu (2013), The Collection 6 MODIS aerosol products over land and ocean, *Atmos. Meas. Tech.*, *6*, 2989–3034.

Mast, M. A., D. H. Campbell, and G. P. Ingersoll (2005), Effects of emission reductions at the Hayden powerplant on precipitation, snowpack, and surface-water chemistry in the Mount Zirkel Wilderness Area, Colorado, 1995–2003, U.S. Geological Survey Scientific Investigations Report 2005–5167, 32 p. [Available at https://store.usgs.gov/yimages/PDF/2.]

McLinden, C. A., V. Fioletov, K. F. Boersma, S. K. Kharol, N. Krotkov, L. Lamsal, P. A. Makar, R. V. Martin, J. P. Veefkind, and K. Yang (2014), Improved satellite retrievals of NO_2 and SO_2 over the Canadian oil sands and comparisons with surface measurements, *Atmos. Chem. Phys.*, *17*(7), 3637–3656.

Merikanto, J., D. V. Spracklen, G. W. Mann, S. J. Pickering, and K. S. Carslaw (2009), Impact of nucleation on global CCN, *Atmos. Chem. Phys.*, *9*(21), 8601–8616.

Metzger, A., B. Verheggen, J. Dommen, J. Duplissy, A. S. H. Prevot, E. Weingartner, I. Riipinen, M. Kulmala, D. V. Spracklen, and K. S. Carslaw (2010), Evidence for the role of organics in aerosol particle formation under atmospheric conditions, *Proc. Natl. Acad. Sci. U.S.A.*, *107*(15), 6646–6651.

Millet, D. B., D. J. Jacob, S. Turquety, R. C. Hudman, S. Wu, A. Fried, J. Walega, B. G. Heikes, D. R. Blake, and H. B. Singh (2006), Formaldehyde distribution over North America: Implications for satellite retrievals of formaldehyde columns and isoprene emission, *J. Geophys. Res.*, *111*, D24S02, doi:10.1029/2005JD006853.

Millet, D. B., D. J. Jacob, K. F. Boersma, T. Fu, T. P. Kurosu, K. Chance, C. L. Heald, and A. Guenther (2008), Spatial distribution of isoprene emissions from North America derived from formaldehyde column measurements by the OMI satellite sensor, *J. Geophys. Res.*, *113*, D02307, doi:10.1029/2007JD008950.

O'Dowd, C. D., P. Aalto, K. Hmeri, M. Kulmala, and T. Hoffmann (2002), Aerosol formation: Atmospheric particles from organic vapours, *Nature*, *416*(6880), 497–498.

OMI Team (2012), Ozone monitoring instrument (OMI) data user's guide, OMI-DUG-5.0. [Available at http://disc.sci.gsfc.nasa.gov/Aura/data-holdings/additional/documentation/README.OMI_DUG.pdf.]

Paasonen, P., T. Nieminen, E. Asmi, H. E. Manninen, T. Petäjä, C. Plass-Dülmer, H. Flentje, W. Birmili, A. Wiedensohler, and U. Horrak (2010), On the roles of sulphuric acid and low-volatility organic vapours in the initial steps of atmospheric new particle formation, *Atmos. Chem. Phys.*, *10*(22), 11,223–11,242.

Paasonen, P., A. Asmi, T. Petäjä, M. K. Kajos, M. Äijälä, H. Junninen, T. Holst, J. P. D. Abbatt, A. Arneth, and W. Birmili (2013), Warming-induced increase in aerosol number concentration likely to moderate climate change, *Nat. Geosci.*, *6*(6), 438–442.

Pierce, J. R., W. R. Leaitch, J. Liggio, D. M. Westervelt, C. D. Wainwright, J. P. D. Abbatt, L. Ahlm, W. Al-Basheer, D. J. Cziczo, and K. L. Hayden (2012), Nucleation and condensational growth to CCN sizes during a sustained pristine biogenic SOA event in a forested mountain valley, *Atmos. Chem. Phys.*, *12*(7), 3147–3163.

Pierce, J. R., D. M. Westervelt, S. A. Atwood, E. A. Barnes, and W. R. Leaitch (2014), New-particle formation, growth and climate-relevant particle production in Egbert, Canada: Analysis from 1 year of size-distribution observations, *Atmos. Chem. Phys.*, *14*(16), 8647–8663.

Pillai, P., A. Khlystov, J. Walker, and V. Aneja (2013), Observation and analysis of particle nucleation at a forest site in southeastern US, *Atmosphere*, *4*(2), 72–93.

Pryor, S. C., A. M. Spaulding, and R. J. Barthelmie (2010), New particle formation in the Midwestern USA: Event characteristics, meteorological context and vertical profiles, *Atmos. Environ.*, *44*(35), 4413–4425, doi:10.1016/j.atmosenv.2010.07.045.

Pryor, S. C., R. J. Barthelmie, L. L. Sørensen, J. G. McGrath, P. Hopke, and T. Petäjä (2011), Spatial and vertical extent of nucleation events in the Midwestern USA: Insights from the Nucleation In ForesTs (NIFTy) experiment, *Atmos. Chem. Phys.*, *11*(4), 1641–1657, doi:10.5194/acp-11-1641-2011.

Pryor, S. C., K. E. Hornsby, and K. A. Novick (2014), Forest canopy interactions with nucleation mode particles, *Atmos. Chem. Phys.*, *14*(21), 11,985–11,996, doi:10.5194/acp-14-11985-2014.

Qi, X. M., A. J. Ding, W. Nie, T. Petäjä, V.-M. Kerminen, E. Herrmann, Y. N. Xie, L. F. Zheng, H. Manninen, and P. Aalto (2015), Aerosol size distribution and new particle formation in the western Yangtze River Delta of China: 2 years of measurements at the SORPES station, *Atmos. Chem. Phys.*, *15*, 12,445–12,464.

Riccobono, F., S. Schobesberger, C. E. Scott, J. Dommen, I. K. Ortega, L. Rondo, J. Almeida, A. Amorim, F. Bianchi, and M. Breitenlechner (2014), Oxidation products of biogenic emissions contribute to nucleation of atmospheric particles, *Science*, *344*(6185), 717–721.

Riipinen, I., T. Yli-Juuti, J. R. Pierce, T. Petäjä, D. R. Worsnop, M. Kulmala, and N. M. Donahue (2012), The contribution of organics to atmospheric nanoparticle growth, *Nat. Geosci.*, *5*(7), 453–458.

Rose, C., K. Sellegri, E. Freney, R. Dupuy, A. Colomb, J.-M. Pichon, M. Ribeiro, T. Bourianne, F. Burnet, and A. Schwarzenboeck (2015), Airborne measurements of new particle formation in the free troposphere above the Mediterranean Sea during the HYMEX campaign, *Atmos. Chem. Phys.*, *15*(17), 10,203–10,218.

Russell, A. R., L. C. Valin, and R. C. Cohen (2012), Trends in OMI NO_2 observations over the United States: Effects of emission control technology and the economic recession, *Atmos. Chem. Phys.*, *12*(24), 12,197–12,209.

Seinfeld, J. H., and S. N. Pandis (2006), *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, Wiley, Hoboken, N. J.

Sipilä, M., T. Berndt, T. Petäjä, D. Brus, J. Vanhanen, F. Stratmann, J. Patokoski, R. L. Mauldin, A.-P. Hyvärinen, and H. Lihavainen (2010), The role of sulfuric acid in atmospheric nucleation, *Science*, *327*(5970), 1243–1246.

Smith, J. N., K. F. Moore, F. L. Eisele, D. Voisin, A. K. Ghimire, H. Sakurai, and P. H. McMurry (2005), Chemical composition of atmospheric nanoparticles during nucleation events in Atlanta, *J. Geophys. Res.*, *110*, D22S03, doi:10.1029/2005JD005912.

Smith, J. N., M. J. Dunn, T. M. VanReken, K. Iida, M. R. Stolzenburg, P. H. McMurry, and L. G. Huey (2008), Chemical composition of atmospheric nanoparticles formed from nucleation in Tecamac, Mexico: Evidence for an important role for organic species in nanoparticle growth, *Geophys. Res. Lett.*, *35*, L04808, doi:10.1029/2007GL032523.

Spracklen, D. V., B. Bonn, and K. S. Carslaw (2008a), Boreal forests, aerosols and the impacts on clouds and climate, *Philos. Trans. R. Soc. Lond. A Math. Phys. Eng. Sci.*, *366*(1885), 4613–4626.

Spracklen, D. V., K. S. Carslaw, M. Kulmala, V. Kerminen, S. Sihto, I. Riipinen, J. Merikanto, G. W. Mann, M. P. Chipperfield, and A. Wiedensohler (2008b), Contribution of particle formation to global cloud condensation nuclei concentrations, *Geophys. Res. Lett.*, *35*, L06808, doi:10.1029/2007GL033038.

Sullivan, R. C., and S. C. Pryor (2016), Dynamic and chemical controls on new particle formation occurrence and characteristics from in situ and satellite-based measurements, *Atmos. Environ.*, *127*, 316–325, doi:10.1016/j.atmosenv.2015.12.050.

Sullivan, R. C., R. C. Levy, and S. C. Pryor (2015), Spatiotemporal coherence of mean and extreme aerosol particle events over eastern North America as observed from satellite, *Atmos. Environ.*, *112*, 126–135, doi:10.1016/j.atmosenv.2015.04.026.

Sundström, A.-M., A. Nikandrova, K. Atlaskina, T. Nieminen, V. Vakkari, L. Laakso, J. P. Beukes, A. Arola, P. G. van Zyl, and M. Josipovic (2015), Characterization of satellite-based proxies for estimating nucleation mode particles over South Africa, *Atmos. Chem. Phys.*, *15*(9), 4983–4996.

Surratt, J. D., S. M. Murphy, J. H. Kroll, N. L. Ng, L. Hildebrandt, A. Sorooshian, R. Szmigielski, R. Vermeylen, W. Maenhaut, and M. Claeys (2006), Chemical composition of secondary organic aerosol formed from the photooxidation of isoprene, *J. Phys. Chem. A*, *110*(31), 9665–9690.

US Environmental Protection Agency (2011), 2011 National Emissions Inventory Downloaded from. [Available at https://www3.epa.gov/ttnchie1/net/.]

Valente, R. J., R. E. Imhoff, R. L. Tanner, J. F. Meagher, P. H. Daum, R. M. Hardesty, R. M. Banta, R. J. Alvarez, R. T. McNider, and N. V. Gillani (1998), Ozone production during an urban air stagnation episode over Nashville, Tennessee, *J. Geophys. Res.*, *103*(D17), 22,555–22,568, doi:10.1029/98JD01641.

van Donkelaar, A., R. V. Martin, and R. J. Park (2006), Estimating ground-level $PM_{2.5}$ using aerosol optical depth determined from satellite remote sensing, *J. Geophys. Res.*, *111*, D21201, doi:10.1029/2005JD006996.

van Donkelaar, A., R. V. Martin, R. J. D. Spurr, E. Drury, L. A. Remer, R. C. Levy, and J. Wang (2013), Optimal estimation for global ground-level fine particulate matter concentrations, *J. Geophys. Res. Atmos.*, *118*, 5621–5636, doi:10.1002/jgrd.50479.

Vinken, G. C. M., K. F. Boersma, A. van Donkelaar, and L. Zhang (2014), Constraints on ship NO_x emissions in Europe using GEOS-Chem and OMI satellite NO_2 observations, *Atmos. Chem. Phys.*, *14*(3), 1353–1369.

Wan, Z. (2008), New refinements and validation of the MODIS land-surface temperature/emissivity products, *Remote Sens. Environ.*, *112*(1), 59–74.

Westervelt, D. M., J. R. Pierce, I. Riipinen, W. Trivitayanurak, A. Hamed, M. Kulmala, A. Laaksonen, S. Decesari, and P. J. Adams (2013), Formation and growth of nucleated particles into cloud condensation nuclei: Model-measurement comparison, *Atmos. Chem. Phys.*, *13*(15), 7645–7663.

Whitburn, S., M. Van Damme, L. Clarisse, S. Bauduin, C. Heald, J. Hadji-Lazaro, D. Hurtmans, M. A. Zondlo, C. Clerbaux, and P.-F. Coheur (2016), A flexible and robust neural network IASI-NH_3 retrieval algorithm, *J. Geophys. Res. Atmos.*, *121*, 6581–6599, doi:10.1002/2016JD024828.

Yu, F., and A. Hallar (2014), Difference in particle formation at a mountaintop location during spring and summer: Implications for the role of sulfuric acid and organics in nucleation, *J. Geophys. Res. Atmos.*, *119*, 12,246–12,255, doi:10.1002/2014JD022136.

Yu, F., and G. Luo (2009), Simulation of particle size distribution with a global aerosol model: Contribution of nucleation to aerosol and CCN number concentrations, *Atmos. Chem. Phys.*, *9*(20), 7691–7710.

Yu, F., et al. (2015), Spring and summer contrast in new particle formation over nine forest areas in North America, *Atmos. Chem. Phys.*, *15*(24), 13,993–14,003, doi:10.5194/acpd-15-21271-2015.

Zhang, Q. I., C. O. Stanier, M. R. Canagaratna, J. T. Jayne, D. R. Worsnop, S. N. Pandis, and J. L. Jimenez (2004a), Insights into the chemistry of new particle formation and growth events in Pittsburgh based on aerosol mass spectrometry, *Environ. Sci. Technol.*, *38*(18), 4797–4809.

Zhang, R., I. Suh, J. Zhao, D. Zhang, E. C. Fortner, X. Tie, L. T. Molina, and M. J. Molina (2004b), Atmospheric new particle formation enhanced by organic acids, *Science*, *304*(5676), 1487–1490.

Zhang, Y., P. H. McMurry, F. Yu, and M. Z. Jacobson (2010), A comparative study of nucleation parameterizations: 1. Examination and evaluation of the formulations, *J. Geophys. Res.*, *115*, D20212, doi:10.1029/2010JD014150.